Artificial DNA: PNA & XNA 5, e29408; June 2014; © 2014 Landes Bioscience

The genetic code

Rewritten, revised, repurposed

Roy D Sleator
Department of Biological Sciences; Cork Institute of Technology; Cork, Ireland

Despite remaining apparently frozen through the millennia, the genetic code is far more flexible than previously believed and can be extended and repurposed with relative ease.

Keywords: synthetic DNA, recoding, unnatural base pairs (UBPs), nonstandard amino acids (NSAAs)

*Correspondence: Roy D Sleator; Email: roy.sleator@cit.ie

Submitted: 05/29/2014

Accepted: 05/30/2014

Published Online: 06/17/2014

Citation: Sleator RD. The genetic code: Rewritten, revised, repurposed. Artificial DNA: PNA & XNA 2014; 5:e29408; PMID: 24937253; http://dx.doi.org/10.4161/adna.29408

Despite the fact that there are more than 100 amino acids observed in nature, only 20 are encoded by the canonical genetic code of 61 sense codons and 3 stop codons. Because sense codons outnumber their encoded amino acids by a ratio of 3:1, the genetic code is redundant; with most amino acids coded for by more than one codon.[1] This degeneracy is well documented, with certain organisms having evolved preferences for specific codon-amino acid combinations.[2] However, despite this inherent flexibility, our natural amino acid repertoire represents less than 20% that which exists in nature; leading Francis Crick to suggest that the code is a "frozen accident."[3] However, several hot papers have emerged in recent years which have led to a significant thaw in this concept of a "frozen" code.

In one of the earliest successful attempts to extend or rewrite the code, Sakamoto and colleagues undertook a process of genetic recoding[4]; forcing specific codons to code for alternative or nonstandard amino acids (NSAAs). Sakamoto's team converted the TAG stop codon in 7 essential *Escherichia coli* genes to TAA; eliminated release factor 1 (RF1; which terminates translation at UAA and UAG) and supplied a tRNA that inserts a glutamine when it encounters UAG. Following proof of concept, with a canonical amino acid, the team repeated the experiment reassigning TAG to the NSAA iodotyrosine. Despite the experiment being a success, with all 7 targeted genes terminating properly, all the remaining genes ending in TAG failed to terminate correctly in the absence of RF1.

Lajoie et al.,[5] overcame this 'read through' limitation by employing an in vivo genome-editing approach[6]; replacing all 321 instances of TAG (the rarest codon in the *E. coli* strain tested) with TAA. The resulting organism described as a genomically recoded organism (GRO), represents a new class of genetically modified organism (GMO) and a potentially important platform for novel drug production. Indeed, in support of the application of GROs as industrial protein production systems, Lajoie et al.,[5] successfully reassigned TAG to a prephosphorlyated serine – a modification found on serines of the recombinant human growth hormone.[7] Furthermore, the GRO exhibited increased resistance to T7 bacteriophage; a highly desirable trait in large scale industrial processes which are otherwise susceptible to phage attack.[8] This observed phage resistance prompted the authors to suggest that genetic recoding may lead to viral protein mistranslation.

In a second paper in the same issue of *Science*, Lajoie et al.,[9] investigated the effect of recoding sense codons; removing all instances of 13 rare codons from 42 highly expressed essential genes (including all 41 essential ribosomal protein-coding genes and *prfB*) across 80 *E. coli* strains. Despite several genome design constraints, growth defects, and the fact that replacement of synonymous codons occasionally did not produce the same effects as the native codon; genome-wide

reassignment of sense codons was at least shown to be possible.

In addition to these laboratory based recoding successes, we are beginning to see more and more variation in the genetic code of natural organisms.[10] Indeed, a recent large scale analysis of stop codon reassignments in the wild revealed far higher recoding rates than previously imagined.[11] Investigating > 1,700 environmental samples (including 750 samples from 17 human body sites), Ivanova et al.,[11] scanned ~5.6 trillion bp of metagenomic data for stop codon reassignment. Contrary to the previously held belief that natural recoding is rare; the authors report a total of 198 Mb of recoded DNA data. Interestingly, the human body despite accounting for only 10% of DNA present represented 51% of all codon reassignments. Furthermore, distinct patterns of stop codon reassignment were observed in all 3 domains of life, with bacteria showing only *opal* reassignments, while extensive *opal* and *amber* reassignments occurred in phages. The observed high rate of recoding among phage suggests that, contrary to the findings of Lajoie et al.,[5,9] phages are not obliged to adapt to the codon usage of their hosts, but rather exploit differences in codon usage to manipulate their hosts.

While genetic recoding or rewriting is still restricted by our existing dependency on the 4 natural nucleotides A, T, G, and C; an alternative approach to extending or revising the code involves the use of unnatural base pairs (UBPs), allowing us to incorporate up to 152 additional non-canonical amino acids. Over the past 15 y, Romesberg and colleagues at the Scripps Research Institute, having synthesized and tested more than 300 artificial nucleotides, developed a class of UBPs, exemplified by d5SICS-dNaM (abbreviated as X and Y), formed between nucleotides bearing hydrophobic nucleobases.[12] Romesberg's group recently proved that it is possible to stably incorporate X and Y into the DNA of actively growing *E. coli*, creating the first organism to stably propagate an expanded genetic alphabet.[13] It is hoped that this expanded DNA alphabet will help to build an expanded translational alphabet; encoding more and more NSAAs, ultimately enabling the synthesis of new and improved proteins.

In addition to being rewritten and revised, perhaps the most innovative use of the genetic code in recent times is its deliberate repurposing as a high capacity storage medium. With a theoretical storage potential of 455 exabytes per gram ssDNA,[14] it is estimated that all of the world's projected 40 ZB of data could be stored in just ~90 g of DNA.[15] Some of the earliest attempts to use DNA as a workable canvas for archival purposes include Joe Davis' Microvenus; a 35 bit coded visual icon representing the external female genitalia.[16] More recently, construction of JCVI-syn1.0, the first bacterial cell to contain a completely synthetic genome, employed "watermarks" to distinguish the synthetic genome from native DNA. These 7,920 bit watermarks contain a web address, the names of the paper's authors and some memorable quotations.[17]

Large scale data storage in DNA was first achieved by Church and colleagues[14] who described the conversion of html-coded data to DNA code using a 1 bit per base encoding (A,C = 0; T,G = 1); allowing the conversion of Church's book *Regenesis* (including 53,426 words, 11 JPG images and 1 JavaScript program) into DNA sequence. In an effort to reduce error and facilitate up-scaling, Goldman et al.[18]., described a modified strategy achieving a storage density of ~2.2 PB/g DNA (Equivalent to ~468,000 DVDs). This modified approach first converts the original file type to binary code (0, 1) which is then converted to a ternary code (0, 1, 2) and in turn to the triplet DNA code. Replacing each trit with 1 of the 3 nucleotides different from the preceding one (*i.e.* A, T, or C, if the preceding one is G) ensures that no homopolymers are generated – significantly reducing high throughput sequencing errors.[19] Based on a fixed string length (data and indexing) of 117 nt, Goldman et al.[18]., suggest that DNA-based storage currently remains feasible even at several orders of magnitude greater than current global data volumes. This, combined with the likely expectation of significantly longer string synthesis as the technology progresses,[20] virtually future proofs DNA as a viable big data storage medium.[21] Furthermore, while the above strategies focus on maintaining DNA in vitro, we have previously postulated that in vivo storage may also be a viable and perhaps even more desirable option.[22]

Therefore, despite remaining apparently frozen through the millennia, advances like those described above, have revealed a code that is far more flexible than we could previously have hoped to believe; a code which we can extend and repurpose with relative ease. While it is difficult to predict future directions in this particular field of synthetic biology,[23] it is clear that several exciting possibilities exist. One prospect is the synthesis of completely novel species; designed and synthesized using the principles described previously,[17,24] yet potentially running multiple genetic codes concurrently. Such hybrid constructs can be thought of as analogous to a computer running multiple operating systems in parallel; each designed for a specific purpose. The native code (consisting of A, T G, and C) would run normal cellular processes, required for growth and reproduction, while the parallel synthetic code (incorporating X and Y) would allow the cell to act as a micro-factory, producing new proteins with novel applications in industry and medicine. Finally, the third partitioned code might contain the manufacturer's instructions, or user's manual, digitally encoded in the DNA.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

RDS is coordinator of the EU FP7 project ClouDx-i.

References

1. Sleator RD. Proteins: form and function. Bioeng Bugs 2012; 3:80-5; PMID:22095055; http://dx.doi.org/10.4161/bbug.18303
2. Johnston C, Douarre PE, Soulimane T, Pletzer D, Weingart H, MacSharry J, Coffey A, Sleator RD, O'Mahony J. Codon optimisation to improve expression of a Mycobacterium avium ssp. paratuberculosis-specific membrane-associated antigen by Lactobacillus salivarius. Pathog Dis 2013; 68:27-38; PMID:23620276; http://dx.doi.org/10.1111/2049-632X.12040
3. Sella G, Ardell DH. The coevolution of genes and genetic codes: Crick's frozen accident revisited. J Mol Evol 2006; 63:297-313; PMID:16838217; http://dx.doi.org/10.1007/s00239-004-0176-7

4. Mukai T, Hayashi A, Iraha F, Sato A, Ohtake K, Yokoyama S, Sakamoto K. Codon reassignment in the Escherichia coli genetic code. Nucleic Acids Res 2010; 38:8188-95; PMID:20702426; http://dx.doi.org/10.1093/nar/gkq707
5. Lajoie MJ, Rovner AJ, Goodman DB, Aerni HR, Haimovich AD, Kuznetsov G, Mercer JA, Wang HH, Carr PA, Mosberg JA, et al. Genomically recoded organisms expand biological functions. Science 2013; 342:357-60; PMID:24136966; http://dx.doi.org/10.1126/science.1241459
6. Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, Kraal L, Tolonen AC, Gianoulis TA, Goodman DB, Reppas NB, et al. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. Science 2011; 333:348-53; PMID:21764749; http://dx.doi.org/10.1126/science.1205822
7. Levarski Z, Soltýsová A, Krahulec J, Stuchlík S, Turna J. High-level expression and purification of recombinant human growth hormone produced in soluble form in Escherichia coli. Protein Expr Purif 2014; (Forthcoming); PMID:24859479; http://dx.doi.org/10.1016/j.pep.2014.05.003
8. Sturino JM, Klaenhammer TR. Engineered bacteriophage-defence systems in bioprocessing. Nat Rev Microbiol 2006; 4:395-404; PMID:16715051; http://dx.doi.org/10.1038/nrmicro1393
9. Lajoie MJ, Kosuri S, Mosberg JA, Gregg CJ, Zhang D, Church GM. Probing the limits of genetic recoding in essential genes. Science 2013; 342:361-3; PMID:24136967; http://dx.doi.org/10.1126/science.1241460
10. Prat L, Heinemann IU, Aerni HR, Rinehart J, O'Donoghue P, Söll D. Carbon source-dependent expansion of the genetic code in bacteria. Proc Natl Acad Sci U S A 2012; 109:21070-5; PMID:23185002; http://dx.doi.org/10.1073/pnas.1218613110
11. Ivanova NN, Schwientek P, Tripp HJ, Rinke C, Pati A, Huntemann M, Visel A, Woyke T, Kyrpides NC, Rubin EM. Stop codon reassignments in the wild. Science 2014; 344:909-13; PMID:24855270; http://dx.doi.org/10.1126/science.1250691
12. Malyshev DA, Dhami K, Quach HT, Lavergne T, Ordoukhanian P, Torkamani A, Romesberg FE. Efficient and sequence-independent replication of DNA containing a third base pair establishes a functional six-letter genetic alphabet. Proc Natl Acad Sci U S A 2012; 109:12005-10; PMID:22773812; http://dx.doi.org/10.1073/pnas.1205176109
13. Malyshev DA, Dhami K, Lavergne T, Chen T, Dai N, Foster JM, Corrêa IR Jr., Romesberg FE. A semi-synthetic organism with an expanded genetic alphabet. Nature 2014; 509:385-8; PMID:24805238; http://dx.doi.org/10.1038/nature13314
14. Church GM, Gao Y, Kosuri S. Next-generation digital information storage in DNA. Science 2012; 337:1628; PMID:22903519; http://dx.doi.org/10.1126/science.1226355
15. O' Driscoll A, Sleator RD. Synthetic DNA: the next generation of big data storage. Bioengineered 2013; 4:123-5; PMID:23514938; http://dx.doi.org/10.4161/bioe.24296
16. Sleator RD. The story of Mycoplasma mycoides JCVI-syn1.0: the forty million dollar microbe. Bioeng Bugs 2010; 1:229-30; PMID:21327053; http://dx.doi.org/10.4161/bbug.1.4.12465
17. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. Science 2010; 329:52-6; PMID:20488990; http://dx.doi.org/10.1126/science.1190719
18. Goldman N, Bertone P, Chen S, Dessimoz C, LeProust EM, Sipos B, Birney E. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. Nature 2013; 494:77-80; PMID:23354052; http://dx.doi.org/10.1038/nature11875
19. Niedringhaus TP, Milanova D, Kerby MB, Snyder MP, Barron AE. Landscape of next-generation sequencing technologies. Anal Chem 2011; 83:4327-41; PMID:21612267; http://dx.doi.org/10.1021/ac2010857
20. Fehér T, Burland V, Pósfai G. In the fast lane: large-scale bacterial genome engineering. J Biotechnol 2012; 160:72-9; PMID:22406111; http://dx.doi.org/10.1016/j.jbiotec.2012.02.012
21. O'Driscoll A, Daugelaite J, Sleator RD. 'Big data', Hadoop and cloud computing in genomics. J Biomed Inform 2013; 46:774-81; PMID:23872175; http://dx.doi.org/10.1016/j.jbi.2013.07.001
22. Sleator RD, O'Driscoll A. Digitizing humanity. Artif DNA PNA XNA 2013; 4:37-8; PMID:23912716; http://dx.doi.org/10.4161/adna.25489
23. Sleator RD. The synthetic biology future. Bioengineered 2014; 5:5; PMID:24561910; http://dx.doi.org/10.4161/bioe.28317
24. Annaluru N, Muller H, Mitchell LA, Ramalingam S, Stracquadanio G, Richardson SM, Dymond JS, Kuang Z, Scheifele LZ, Cooper EM, et al. Total synthesis of a functional designer eukaryotic chromosome. Science 2014; 344:55-8; PMID:24674868; http://dx.doi.org/10.1126/science.1249252